

Segmentation and feature extraction of cervical spine x-ray images

L. Rodney Long^{*}, George R. Thoma

National Library of Medicine, Bethesda, MD 20894

ABSTRACT

As part of an R&D project in mixed text/image database design, the National Library of Medicine has archived a collection of 17,000 digitized x-ray images of the cervical and lumbar spine which were collected as part of the second National Health and Nutrition Examination Survey (NHANES II). To make this image data available and usable to a wide audience, we are investigating techniques for indexing the image content by automated or semi-automated means. Indexing of the images by features of interest to researchers in spine disease and structure requires effective segmentation of the vertebral anatomy. This paper describes work in progress toward this segmentation of the cervical spine images into anatomical components of interest, including anatomical landmarks for vertebral location, and segmentation and identification of individual vertebrae. Our work includes developing a reliable method for automatically fixing an anatomy-based coordinate system in the images, and work to adaptively threshold the images, using methods previously applied by researchers in cardioangiography. We describe the motivation for our work and present our current results in both areas.

Keywords: medical x-ray image, segmentation, image processing, radiograph, cervical spine, NHANES, image database, NLM

1. INTRODUCTION

At the Lister Hill National Center for Biomedical Communications, a research and development part of the National Library of Medicine, we are building a biomedical information resource consisting of digitized x-ray images and associated national health survey data. We have described the architecture and characteristics of this system in previous papers¹⁻². This resource is called the *Web-based Medical Information Retrieval System*, or WebMIRS. This paper discusses our current work toward the long-range goal of directly exploiting the image contents in our system. In a future WebMIRS system, we envision that we will have in our databases not only *text and raw image data*, but *quantitative anatomical feature information* that is derived directly from the images. We further envision the system to have capability to retrieve images based on image characteristics, either alone or in conjunction with text descriptions associated with the images.

The current version of WebMIRS lets the user create a query using a Graphical User Interface and get the results displayed as either a table of database text records or as a display of images and related text. **Figure 1** shows a sample output screen from WebMIRS. WebMIRS is an Internet database application, written as a Java applet, and usable by means of a standard Web browser.

As detailed in the WebMIRS papers referenced above, our archive consists of records for two databases, corresponding to the survey data collected in the second and third National Health and Nutrition Examination Surveys (NHANES). For the NHANES II survey, the records contain information for approximately 20,000 survey participants. Each record contains about two thousand data points, including demographic information, answers to health questionnaires, anthropometric information, and the results of a physician's examination. In addition, many of the participants were x-rayed: approximately 10,000 cervical spine and 7,000 lumbar spine x-rays were collected. The WebMIRS system makes the health data available for query in a relational database system, and, at the user's option, also returns the image data for display. Image results returned are raw images only; no quantitative or even descriptive information about the images themselves is stored in the database.

The images in this survey were collected primarily for the study of osteoarthritis and degenerative disc disease. Biomedical features of interest to researchers in these areas have been identified by two workshops conducted at NIH, and consist of anterior osteophytes, disc space narrowing, and subluxation. A database containing quantitative and descriptive image information that allows intelligent search for images that show varying degrees of these features is expected to be of research use for the osteoarthritis and related communities. In addition, published research work in the field of vertebral morphometry suggests that a database with quantitative measures of vertebral dimensions may be useful for purposes such as studies in occurrence of spinal fracture and estimation of normative values for vertebral size.

In order to derive such information directly from the images, detailed analysis and measurement of each image is required. Due to the prohibitive cost of carrying out such a process on 17,000 images, we are conducting research into methods for automating or semi-automating the process. In this paper, we describe our initial work toward this goal, which consists of (1) developing an algorithm to find basic landmarks in the images by fixing a coordinate system in the anatomy itself, and (2) work to implement an adaptive grayscale thresholding method for isolating the anatomical components of the images.

2. APPROACH

In our current work we are approaching the segmentation as a problem in (1) obtaining basic orientation and landmark identification in the large-scale image structures and (2) using this information to locate the vertebral region to segment and identify the individual vertebrae. The work reported in this paper applies to the first step of basic image orientation (establishing an anatomy-based coordinate system) and to part of the process of segmenting the vertebrae in the second step. We have confined our work to this point to the cervical spine images.

The purpose of establishing a coordinate system fixed in the anatomy of the figure in the image is to get a basic orientation in image content: to find a fundamental landmark and frame of reference which is tied to the image content itself, rather than only to the "frame" of the image; in so doing, this anatomy-based system is expected to compensate for much of the variation in positioning of the human subject within the image frame. To achieve this basic orientation in the image, we define an orthogonal coordinate system in terms of features usually visible in human skeletal anatomy as seen in the profile projections given by cervical spine radiographs. We then seek an automated method of locating the axes of this system in a given image. We expect that such a known coordinate system within a given image will be extremely useful in navigating through and identifying the individual vertebrae. The algorithm we are developing is heuristic in nature, but based on geometric characteristics of the skull and spine expected from simple anatomical assumptions. The reason for this step is further explained in Section 3.

One of the requirements of our algorithm for fixing this coordinate system in the images is good grayscale thresholding, in particular on the borders of the skull, as will be seen in Section 3. For this reason, and also to investigate grayscale thresholding as a method for segmenting the individual vertebrae, we are implementing and evaluating an adaptive technique used by Chow and Kanetko³⁻⁴ on angiograph images, described in Section 4. We intend to adapt and apply the Chow-Kanetko algorithm to our image domain of interest.

The digitized cervical spine images are 1463x1755 12-bit grayscale images. For ease of manipulation and computation, we are experimenting with using reduced resolution versions of these images for basic orientation and gross level segmentation. However, we expect that fine level segmentation will be done on identified regions of interest within the larger images at full spatial resolution. The images in the current work are 365x438 8-bit grayscale, essentially one-fourth the spatial resolution of the originals.

Our algorithms have all been coded in the MATLAB language (as "m-files") and tested under MATLAB 5.2, using the Image Processing and the Numerical Algorithms Group (NAG) Toolboxes. Development has been on Windows NT Pentium class computers.

In the following discussion, we follow MATLAB "row/column" conventions for referring to pixels within our images. Rows are indexed 1-N *down* the image, and columns, 1-M *across* the image, from left to right.

3. ANATOMY-BASED COORDINATE SYSTEM

Among the gross level images features expected to be *least* difficult to reliably locate in the images by automated methods are the skull posterior/background boundary region and, for most of the images, the jaw/background boundary region. This suggests that one axis of a coordinate system may be defined as a line W tangent to the “skull base”, passing through the most prominent protrusion on the underside of the jaw and the most prominent protrusion in the skull posterior region. Referring to a chart of human skeletal anatomy indicates that such a line passes through the vertebral column at the nominal location of the C2 vertebra. It also may be observed that, on either side of this intersection with the vertebral column, W passes through a boundary region containing no bony matter. We use these characteristics to complete our coordinate system definition, as follows: we define a second axis V as a line orthogonal to W , and passing through point O on W , such that O is the center of mass for the points in the spinal column which intersect W —it is expected to lie nominally “in the middle” of the spinal column. O is then the origin of the coordinate system. These W/V axes define the conceptual system we want to approximate with the image data.

The algorithm we are developing (“WVLOC”) attempts to establish this coordinate system within a given image by automatically locating the required skull and jaw regions, calculating search curves in these regions for the points defining the W axis, then searching along the W axis for a point O meeting the requirements as an intersection point for the V axis (as indicated in the previous paragraph). From this point we will refer to the process of calculating this coordinate system as “locating” or “fixing” the coordinate system in the image.

The WVLOC algorithm may be summarized as follows:

Given image I ,

- Remove borders from the image
- Apply grayscale thresholding to image I that preserves the essential shape of the skull posterior region and the jaw region.
- Remove small isolated “clutter” in the image left from the thresholding.
- Apply edge detection to the thresholded image.
- Identify the edge curves in the jaw region and the skull posterior region.
- For the edge curve in the jaw region, apply curve slope criteria to each point to identify a subregion of the curve where the slopes lie in an acceptable range for the jaw point; likewise, identify a subregion of the edge curve in the skull posterior region where the slopes lie in an acceptable range for the skull posterior point.
- For each point in the jaw edge curve acceptable subregion, and for each point in the skull posterior edge curve acceptable subregion, find the pair of points (W_J, W_S) which have the closest matching slopes. Take these points as defining the axis W .
- Search along W for a point satisfying the requirements for the point O , origin of the coordinate system, thereby establishing the V axis orthogonal to W .

Details of these steps are given below.

Border removal. Most of the images show extremely bright grayscale values along one or more edges, due to illumination from the scanner leaking around the edge of the x-ray film. These effects appear to be effectively removed by simply deleting a range of pixels around each edge, using image-independent values determined by inspection of a number of the images.

Grayscale thresholding. For good location of the W axis, it is critical that the image be thresholded to preserve the essential shape characteristics of the skull posterior and the underside of the jaw. For the initial work we have used a global threshold computed as the image I mean grayscale value. This was seen to be inadequate in some cases, and we later provide an example of how a failure of the algorithm can be corrected with better thresholding. The need for more effective thresholding has led to the work described in Section 4.

Clutter removal. After thresholding, the images frequently contained a few isolated groups of small numbers of white pixels. We used morphological erosion with a 2x2 kernel identically one (all 1's), which removed most of these.

Edge detection. To the thresholded (now binary) de-cluttered image, we applied a 3x3 maximum gradient edge detector, which uses the Sobel filter.

Figure 2(A) shows an example cervical spine image; **Figure 2(B)** shows the image after the above operations of border removal, thresholding, clutter removal, and edge detection have been done, effectively separating the foreground and background of the image.

Identify the edge curves in the jaw and skull posterior regions. For the skull posterior edge curve detection we begin at the upper right of the edge image and search by scanning rows right to left for edge points; for the jaw edge curve detection, we begin at the lower left of the edge image and search by scanning columns left to right for edge points.

Identify acceptable subregions on the jaw and skull posterior edge curves. For the jaw curve we identify an acceptable subregion for W_j on this curve, by requiring the tangent lines to satisfy slope criteria acceptable for detection of the underside of the jaw. **Figure 3** illustrates the angle criteria applied in this step. We currently require a point q to satisfy $\theta_{tj}(q) > 0$ degrees to accept q as the begin point for the acceptable subregion; we search the jaw curve left-to-right for acceptable points; to end the search we apply a stopping slope criteria intended to detect the downward sloping tangents when the neck region begins. For the skull curve, we apply similar criteria to identify an acceptable subregion for W_s by searching for point p satisfying $\theta_{ts}(p) < -20$ degrees to begin the search; we stop the search when the slope passes a threshold intended to detect the back of neck region. Figure 4 shows slope values for the acceptable subregion of the skull curve, and Figure 5, slope values for the acceptable subregion of the jaw curve; both apply to the image in **Figure 2**. **Figure 5** illustrates the effect of the thresholding step on the algorithm. By visual inspection of the original image of **Figure 2(A)** alone, one might expect the jaw curve to have approximately constant positive values; however, the algorithm operates on the edge curve detected on the thresholded image, and this curve, as shown in **Figure 2(B)** shows a departure from the true jaw contour in the grayscale “cloud” under the jaw and to the front of the throat.

Search the acceptable subregions for the points defining W . We apply an exhaustive search for the two points (W_j^*, W_s^*) , lying in the acceptable subregions for the jaw and skull, respectively, which have the “best matching slopes”. Specifically, we want (W_j^*, W_s^*) so that $\text{abs}(\theta_{tj}(q) - (180 + \theta_{ts}(p)))$ is minimal among all the points q in the jaw acceptable subregion and p in the skull acceptable subregion, respectively. We then take the defining points (W_j, W_s) for the W axis as $(W_j, W_s) = (W_j^*, W_s^*)$. **Figure 2(C)** shows the W axis determined by the algorithm for one image.

Search along W for a point meeting requirements for the coordinate system origin O . In order to determine the point O , we search along the axis W for the “point of maximum grayscale value”, which is expected to lie at the center of mass of the segment of W which intersects the spine; this segment is expected to be bounded by dark grayscale regions which mark the spine boundaries. In order to make the algorithm less sensitive to random variations in the individual point values along W , we use multiple grayscale values at each point on W , as follows: let $W(s) = (r(s), c(s))$ be the representation of the axis W in terms of parameter s , where $r(s)$, $c(s)$ are the row and column coordinates for W at point s . Then each point s , we calculate a line segment $L(t;s) = (r(t), c(t))$, $t=1, \dots, 2^*m + 1$, where $L(;s)$ is orthogonal to W at point s , and L is of length m on each side of W . Then we calculate our grayscale test function as $G(s) = \text{sum}(g(L(t;s)))$, $t=1, \dots, 2^*m+1$, where $g(L(t;s))$ is the image grayscale value at $L(t;s) = (r(t), c(t))$, where t is the parameterization for $L(;s)$. That is, at point s on W , $G(s)$ is the sum of the grayscale values on a orthogonal line segment of length 2^*m , which intersects W at s and is symmetric about W . The geometry for this scheme is shown in **Figure 6**.

Figure 7 shows an actual $G(s)$ curve along the W axis of the image in Figure 2. The curve has been smoothed by replacing each point on the original $G(s)$ with the mean value of k neighboring points ($k=15$ in the current implementation) in order to reduce the number of local extrema in the curve prior to the next step. The region corresponding to the spine area is marked in the Figure, and the dark boundaries around the spine are indicated as well. The small local maximum which follows the second spine boundary (the second “B” in the figure) corresponds to the spinous process on the C1 vertebra, which the W axis intersects for this image (see **Figure 2(C)**).

The detection of the spine boundaries B of **Figure 7** is done by (1) computing the approximate derivative curve $G'(s)$ for $G(s)$, (2) finding the zero-crossings for $G'(s)$, and (3) finding the first two zero-crossings which correspond to local minima. The computation of $G'(s)$ is done by using simple forward differences on the curve $G(s)$: $G'(s) = G(s+d) - G(s)$. (Currently,

d=5 in the algorithm.) **Figure 8** shows the zero-crossings curve for the $G'(s)$ corresponding to **Figure 7**. The zero-crossings that correspond to the spine boundaries are marked “B”.

Figure 9 illustrates the results of the algorithm applied to four other images where it successfully located the W/V axis and shows the level of accuracy, which may be expected.

Figure 10 illustrates failure of the algorithm due to poor thresholding, and how the results which may be improved by more sophisticated thresholding. In **Figure 10(A)** thresholding by global mean grayscale value results in edge curve where the true jaw contour is very badly approximated by the detected edge in the jaw region. The result is a correspondingly bad location of the W_J point for the W axis, as shown in **Figure 10(B)**. To investigate the result of better thresholding, we manually sampled the image grayscale values in the region below but near to the jaw and near to the posterior of the skull, and used the maximum of these values to manually threshold the image. The resulting edge curves, shown in **Figure 10(C)** capture enough of the jaw contour to allow the algorithm to succeed, as shown in **Figure 10(D)**.

Significance of the coordinate system. We expect that being able to reliably locate the W/V coordinate system that we have defined will be an important step in begin able to locate individual vertebrae within the images, since vertebrae locations have a direct relationship to the W/V system through human anatomy. Variability in locations of vertebrae across the images precludes the possibility of accurately locating vertebrae by simply collecting statistics on manually acquired locations from a large number of images. This variance in location of a particular vertebra from image to image may be modeled as: total variance = variance due to differences in subject positioning relative to the frame of the image + variance due to differences in subject head and neck orientation + variance due to differences in subject anatomy (differences in geometry and sizes of vertebrae from subject to subject) + random error variance. For example, manually measuring the location of the center of gravity for the C3 vertebra CGC3 for a large number of images and then using this value as an a priori estimate for CGC3 is impractical, because of the large variance observed, which may be largely attributed to the first component of the total variance—the variance due to differences in subject positioning relative to the frame of the image. However, measuring CGC3 relative to the W/V system removes this component, since the W/V system is fixed on the subject and therefore not sensitive to shifts in the subject relative to the image frame. As an illustration, we manually obtained CGC3 in image frame coordinates (row, col) for the five images shown in Figures 2 and 3, and transformed them to W/V coordinates, using the W/V system automatically returned by the algorithm. The image frame coordinates had standard deviations of (25.5,51.1) respectively, while the W/V coordinates had standard deviations of (9.2,16.7). **Figure 11** shows the CGC3 values for both the image frame coordinates and the W/V coordinates, plotted relative to their means (which are coincident at the point (0,0) in the figure.) The figure suggests that using the mean value of measured CGC3 W/V coordinates from sample images might be a good a priori estimate for CGC3 in a new image, since the CGC3 values cluster more closely around this mean, as compared to attempting to carry out the same process using coordinates fixed in the image frame.

4. FUTURE DIRECTION: ADAPTIVE THRESHOLDING

The above results indicate that automatic location of an anatomy-based coordinate system may be feasible for the cervical spine images. The WVLOC algorithm failure illustrated by **Figure 10** shows that effective grayscale thresholding beyond simple methods is required. Beyond the need for better thresholding for finding jaw and skull contours, such thresholding may prove invaluable in segmenting the finer detail of the vertebrae. Toward this end, we are investigating the adaptive method applied by Chow and Kanetko to angiography

Summary of the Chow-Kanetko algorithm. In the algorithm the image is divided into overlapping subimages. The grayscale distribution for each subimage is then classified as unimodal or bimodal; for the bimodal subimages, the grayscale distribution is (parametrically) modeled as a mixed distribution of two Gaussian densities: one density models the bright “foreground” pixels, and the other, the dark, “background” pixels. The parameters of this model are estimated by minimizing the mean square error between the grayscale values predicted by the model and the grayscale values observed in the subimage. The threshold for the subimage is computed as a function of these estimated parameters. The threshold is optimal in the minimum probability of error sense, i.e., using this threshold to classify pixels as foreground or background minimizes the error that a pixel is misclassified, according to the mixed distribution model being used. Thresholds for subimages without bimodal distributions are found by interpolation. Once all subimages have an assigned threshold, each point in the image is assigned a threshold by interpolation from the subimage thresholds. A detailed description of the algorithm is given in [3] and summarized in [5].

5. CONCLUSION

We have implemented a method for the automatic location of an anatomy-based W/V coordinate system in digitized cervical spine x-ray images. Work in progress on a small number of images has successfully located the W/V system with visually good approximations; some failures, which have been observed in the algorithm, are due to the use of grayscale thresholding, which fails to capture contours on the jaw part of the image. Implementation of an adaptive thresholding technique previously used in angiography is planned as a possible remedy for these failures. We expect to integrate this algorithm into our work and to perform an evaluation of the coordinate-system location with a significant number of images.

REFERENCES

1. Long LR, Goh G-H, Neve L, Thoma GR. Architecture for biomedical information delivery on the World Wide Web, *Proceedings of SPIE Multimedia Storage and Archiving Systems II*, SPIE vol. 3229, Dallas, TX, November 3-4, 1997, pp. 160-169.
2. Long LR, Pillemer SR, Lawrence RC, Goh G-H, Neve L, Thoma GR. WebMIRS: Web-based Medical Information Retrieval System. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases IV*, San Jose, CA, January 28-30, 1998, SPIE vol. 3312, pp. 392-403.
3. Chow CK, Kaneko T. Automated Boundary Detection of the Left Ventricle from Cineangiograms. *Computers and Biomedical Research*. Vol. 5, Num. 4. August 1972, pp. 388-410.
4. Chow CK, Kaneko T. Boundary detection and volume determination of the left ventricle from a cineangiogram. *Computers in Biology and Medicine*. Vol. 3, 1973, pp. 13-26.
5. Gonzalez RC, Woods RE. Digital Image Processing. Addison-Wesley, New York, 1992.



Figure 1. WebMIRS output screen, resulting from a database query to find all records for people over 60 who had low back pain for at least a two-week period. The text at bottom gives field values for one of the people whose image is shown at the top of the screen.

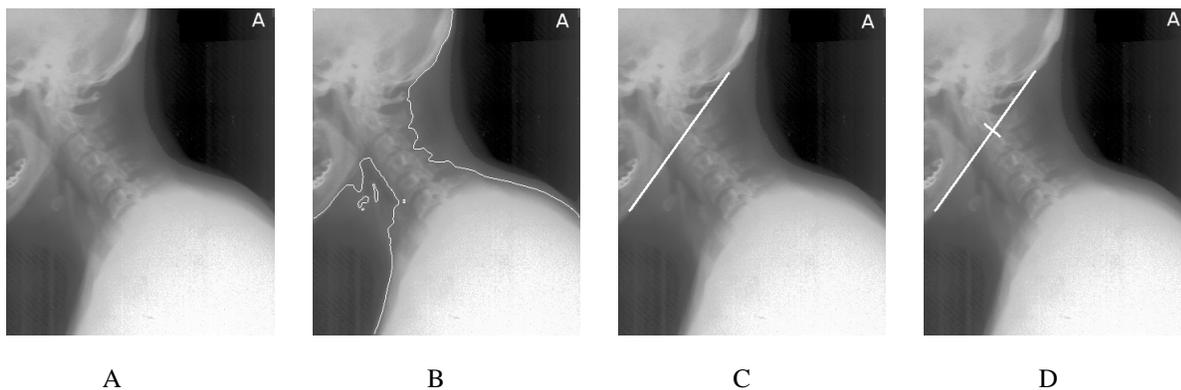


Figure 2. (A) Original cervical spine image with borders removed. (B) Foreground/background separation. (C) W axis fixed. (D) W and V axes fixed.

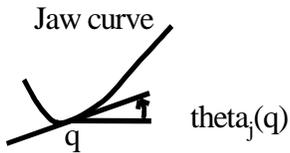
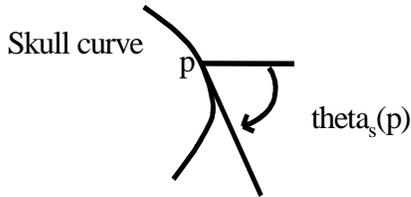


Figure 3. $\theta_{s}(p)$: slope of the skull curve at point p ; $\theta_{j}(q)$: slope of the jaw curve at point q .

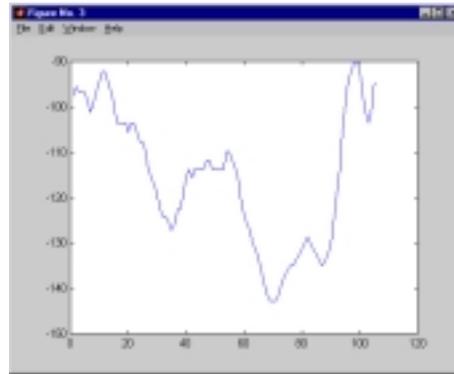


Figure 4. The skull curve $\theta_{s}(p)$ for the acceptable subregion for W_s for the image of Figure 2. Vertical axis is degrees.

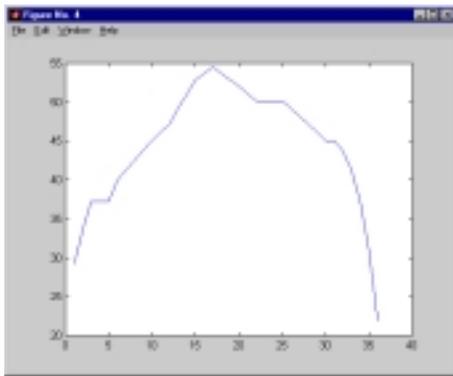


Figure 5. The jaw curve $\theta_{j}(q)$ for the acceptable subregion for W_j for the image of Figure 2. Vertical axis is degrees. Note the downturn in the slope values, which would not be expected from viewing the image itself, but can be understood by viewing the edge detected by thresholding (see Figure 2(B)).

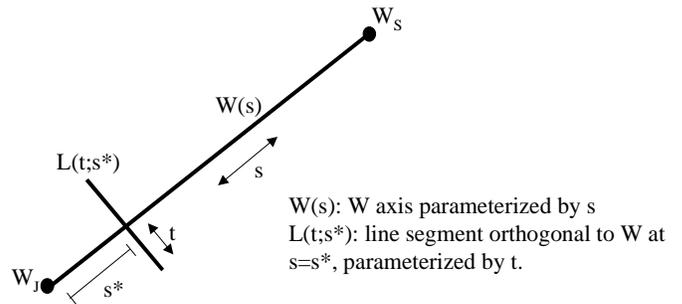


Figure 6. Geometry for computing the gray scale values $G(s^*)$ along the W axis. At each s^* , $G(s^*)$ is computed as the sum of the gray scale values along the orthogonal line segment L : $G(s^*) = \text{Sum}(g(L(t;s^*)))$, where $g()$ is grayscale value, and t ranges over all points on L .

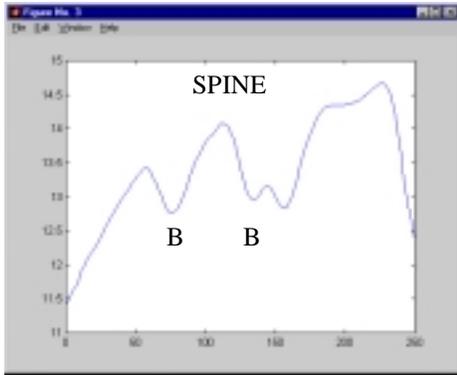


Figure 7. The curve $G(s)$ for the W axis of **Figure 2(C)**, smoothed by neighborhood averaging to reduce local extrema. The spine region lies between the two boundaries marked by the local minima B . The local maximum following the second B corresponds to the spinous process on $C1$, which the W axis intersects for the image.

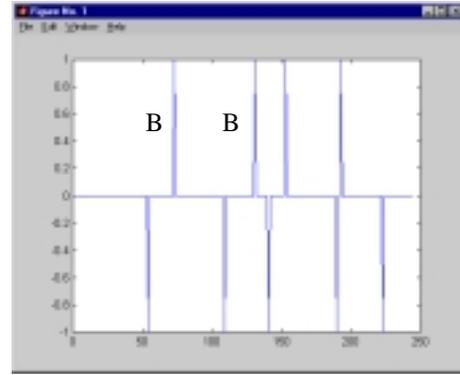


Figure 8. Zero-crossings for $G'(s)$, which approximates the derivative of curve $G(s)$ of Figure 6. The zero-crossings indicate local extrema in $G(s)$. A value of $+1$ indicates that $G(s)$ was increasing at this point (local minimum); a value of -1 indicates that $G(s)$ was decreasing (local maximum). The spine boundaries B were detected as the first two local minima, i.e. points where the grayscale values given by $G(s)$ were at local minimum values.

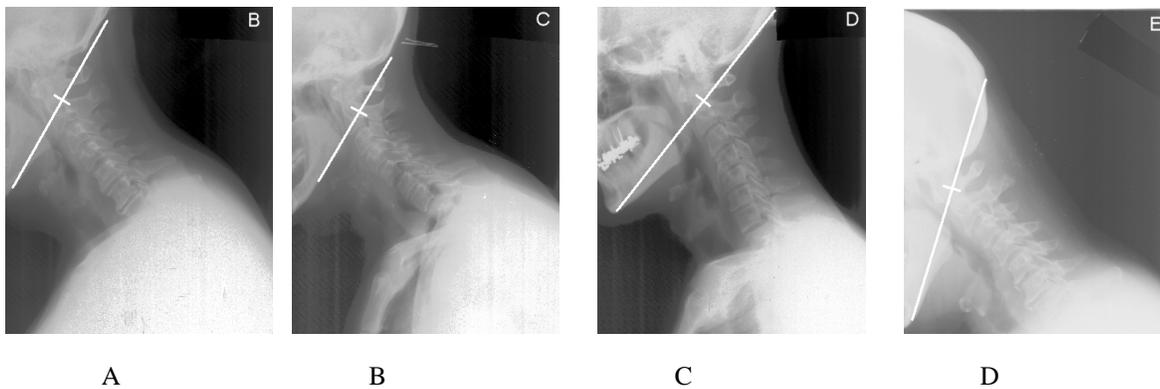


Figure 9. Sample results of the algorithm illustrating the type of accuracy that may be expected in successful cases. Note that in (A) and (B) the W axis appears close to tangent to the skull, while in (C) and (D) small non-tangency can be observed in the jaw and skull regions, respectively. In (A-C) the origin is very near the $C1/C2$ area, while in (D) the origin is on $C1$.

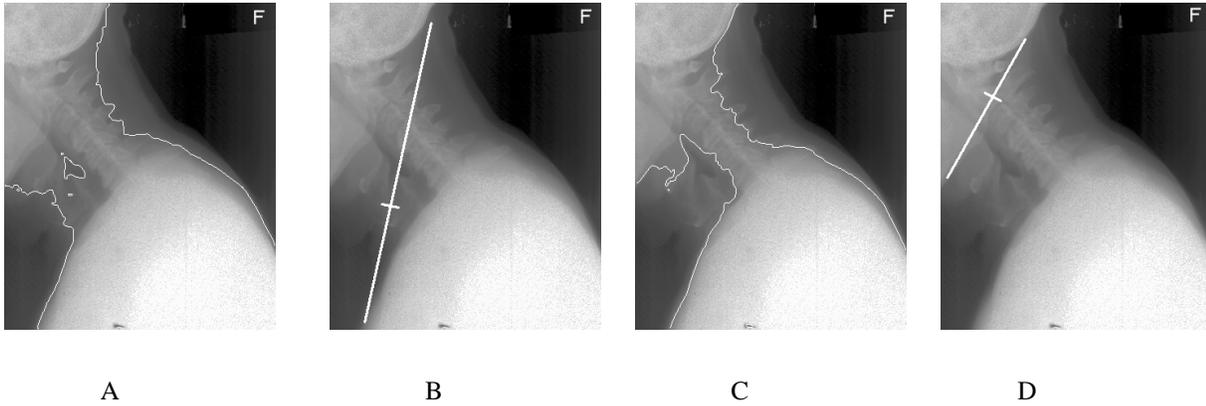


Figure 10. Illustration of effects of poor thresholding. (A) The jaw edge curve for this image is very inaccurate in following the jaw contour when global mean grayscale thresholding is used. (B) Results of the algorithm for (A). (C) The same image was manually thresholded after manually sampling grayscale values below the jaw in near the skull posterior. Now, a fraction of the edge curve follows the jaw contour more reasonably. (D) Results of the algorithm on (C).

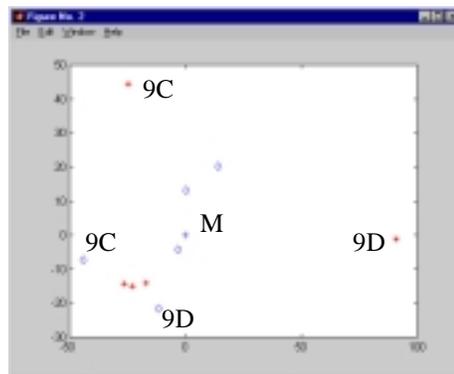


Figure 11. Locations of center of gravity for C3 (CGC3) were manually measured for the five images of Figures 2 and 3. The *'s show these locations relative to the mean (M). The diamonds show the same locations, transformed to the W/V system. If M is used as an estimator for CGC3, note the large error which occurs for 9C and 9D if image frame coordinates are used; the errors are much smaller for W/V coordinates for the same images.